

RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.

Application Serial Number: 10/763,712C
Source: 1FW/b
Date Processed by STIC: 10/23/06

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:50

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

3 <110> APPLICANT: Solazyme, Inc.
 4 Dillon, Harrison F.
 6 <120> TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen
 7 Production
 9 <130> FILE REFERENCE: H2042101-CIP
 11 <140> CURRENT APPLICATION NUMBER: US 10/763,712C
 12 <141> CURRENT FILING DATE: 2004-01-21
 14 <150> PRIOR APPLICATION NUMBER: US 10/287,750
 15 <151> PRIOR FILING DATE: 2002-11-04
 17 <150> PRIOR APPLICATION NUMBER: US 10/411,910
 18 <151> PRIOR FILING DATE: 2003-04-12
 20 <150> PRIOR APPLICATION NUMBER: US 60/500,032
 21 <151> PRIOR FILING DATE: 2003-09-03
 23 <160> NUMBER OF SEQ ID NOS: 199
 25 <170> SOFTWARE: PatentIn version 3.2
 27 <210> SEQ ID NO: 1
 28 <211> LENGTH: 574
 29 <212> TYPE: PRT
 30 <213> ORGANISM: Clostriduim pasteuranum
 32 <400> SEQUENCE: 1
 34 Met Lys Thr Ile Ile Ile Asn Gly Val Gln Phe Asn Thr Asp Glu Asp
 35 1 5 10 15
 38 Thr Thr Ile Leu Lys Phe Ala Arg Asp Asn Asn Ile Asp Ile Ser Ala
 39 20 25 30
 42 Leu Cys Phe Leu Asn Asn Cys Asn Asn Asp Ile Asn Lys Cys Glu Ile
 43 35 40 45
 46 Cys Thr Val Glu Val Glu Gly Thr Gly Leu Val Thr Ala Cys Asp Thr
 47 50 55 60
 50 Leu Ile Glu Asp Gly Met Ile Ile Asn Thr Asn Ser Asp Ala Val Asn
 51 65 70 75 80
 54 Glu Lys Ile Lys Ser Arg Ile Ser Gln Leu Leu Asp Ile His Glu Phe
 55 85 90 95
 58 Lys Cys Gly Pro Cys Asn Arg Arg Glu Asn Cys Glu Phe Leu Lys Leu
 59 100 105 110
 62 Val Ile Lys Tyr Lys Ala Arg Ala Ser Lys Pro Phe Leu Pro Lys Asp
 63 115 120 125
 66 Lys Thr Glu Tyr Val Asp Glu Arg Ser Lys Ser Leu Thr Val Asp Arg
 67 130 135 140
 70 Thr Lys Cys Leu Leu Cys Gly Arg Cys Val Asn Ala Cys Gly Lys Asn
 71 145 150 155 160
 74 Thr Glu Thr Tyr Ala Met Lys Phe Leu Asn Lys Asn Gly Lys Thr Ile
 75 165 170 175
 78 Ile Gly Ala Glu Asp Glu Lys Cys Phe Asp Asp Thr Asn Cys Leu Leu

see p. 6

RAW SEQUENCE LISTING

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:50

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

```

79          180          185          190
82 Cys Gly Gln Cys Ile Ile Ala Cys Pro Val Ala Ala Leu Ser Glu Lys
83          195          200          205
86 Ser His Met Asp Arg Val Lys Asn Ala Leu Asn Ala Pro Glu Lys His
87          210          215          220
90 Val Ile Val Ala Met Ala Pro Ser Val Arg Ala Ser Ile Gly Glu Leu
91 225          230          235          240
94 Phe Asn Met Gly Phe Gly Val Asp Val Thr Gly Lys Ile Tyr Thr Ala
95          245          250          255
98 Leu Arg Gln Leu Gly Phe Asp Lys Ile Phe Asp Ile Asn Phe Gly Ala
99          260          265          270
102 Asp Met Thr Ile Met Glu Glu Ala Thr Glu Leu Val Gln Arg Ile Glu
103          275          280          285
106 Asn Asn Gly Pro Phe Pro Met Phe Thr Ser Cys Cys Pro Gly Trp Val
107          290          295          300
110 Arg Gln Ala Glu Asn Tyr Tyr Pro Glu Leu Leu Asn Asn Leu Ser Ser
111 305          310          315          320
114 Ala Lys Ser Pro Gln Gln Ile Phe Gly Thr Ala Ser Lys Thr Tyr Tyr
115          325          330          335
118 Pro Ser Ile Ser Gly Leu Asp Pro Lys Asn Val Phe Thr Val Thr Val
119          340          345          350
122 Met Pro Cys Thr Ser Lys Lys Phe Glu Ala Asp Arg Pro Gln Met Glu
123          355          360          365
126 Lys Asp Gly Leu Arg Asp Ile Asp Ala Val Ile Thr Thr Arg Glu Leu
127          370          375          380
130 Ala Lys Met Ile Lys Asp Ala Lys Ile Pro Phe Ala Lys Leu Glu Asp
131 385          390          395          400
134 Ser Glu Ala Asp Pro Ala Met Gly Glu Tyr Ser Gly Ala Gly Ala Ile
135          405          410          415
138 Phe Gly Ala Thr Gly Gly Val Met Glu Ala Ala Leu Arg Ser Ala Lys
139          420          425          430
142 Asp Phe Ala Glu Asn Ala Glu Leu Glu Asp Ile Glu Tyr Lys Gln Val
143          435          440          445
146 Arg Gly Leu Asn Gly Ile Lys Glu Ala Glu Val Glu Ile Asn Asn Asn
147          450          455          460
150 Lys Tyr Asn Val Ala Val Ile Asn Gly Ala Ser Asn Leu Phe Lys Phe
151 465          470          475          480
154 Met Lys Ser Gly Met Ile Asn Glu Lys Gln Tyr His Phe Ile Glu Val
155          485          490          495
158 Met Ala Cys His Gly Gly Cys Val Asn Gly Gly Gly Gln Pro His Val
159          500          505          510
162 Asn Pro Lys Asp Leu Glu Lys Val Asp Ile Lys Lys Val Arg Ala Ser
163          515          520          525
166 Val Leu Tyr Asn Gln Asp Glu His Leu Ser Lys Arg Lys Ser His Glu
167          530          535          540
170 Asn Thr Ala Leu Val Lys Met Tyr Gln Asn Tyr Phe Gly Lys Pro Gly
171 545          550          555          560
174 Glu Gly Arg Ala His Glu Ile Leu His Phe Lys Tyr Lys Lys
175          565          570

```

RAW SEQUENCE LISTING

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:50

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

```

178 <210> SEQ ID NO: 2
179 <211> LENGTH: 421
180 <212> TYPE: PRT
181 <213> ORGANISM: Desulfovibrio vulgaris
183 <400> SEQUENCE: 2
185 Met Ser Arg Thr Val Met Glu Arg Ile Glu Tyr Glu Met His Thr Pro
186 1 5 10 15
189 Asp Pro Lys Ala Asp Pro Asp Lys Leu His Phe Val Gln Ile Asp Glu
190 20 25 30
193 Ala Lys Cys Ile Gly Cys Asp Thr Cys Ser Gln Tyr Cys Pro Thr Ala
194 35 40 45
197 Ala Ile Phe Gly Glu Met Gly Glu Pro His Ser Ile Pro His Ile Glu
198 50 55 60
201 Ala Cys Ile Asn Cys Gly Gln Cys Leu Thr His Cys Pro Glu Asn Ala
202 65 70 75 80
205 Ile Tyr Glu Ala Gln Ser Trp Val Pro Glu Val Glu Lys Lys Leu Lys
206 85 90 95
209 Asp Gly Lys Val Lys Cys Ile Ala Met Pro Ala Pro Ala Val Arg Tyr
210 100 105 110
213 Ala Leu Gly Asp Ala Phe Gly Met Pro Val Gly Ser Val Thr Thr Gly
214 115 120 125
217 Lys Met Leu Ala Ala Leu Gln Lys Leu Gly Phe Ala His Cys Trp Asp
218 130 135 140
221 Thr Glu Phe Thr Ala Asp Val Thr Ile Trp Glu Glu Gly Ser Glu Phe
222 145 150 155 160
225 Val Glu Arg Leu Thr Lys Lys Ser Asp Met Pro Leu Pro Gln Phe Thr
226 165 170 175
229 Ser Cys Cys Pro Gly Trp Gln Lys Tyr Ala Glu Thr Tyr Tyr Pro Glu
230 180 185 190
233 Leu Leu Pro His Phe Ser Thr Cys Lys Ser Pro Ile Gly Met Asn Gly
234 195 200 205
237 Ala Leu Ala Lys Thr Tyr Gly Ala Glu Arg Met Lys Tyr Asp Pro Lys
238 210 215 220
241 Gln Val Tyr Thr Val Ser Ile Met Pro Cys Ile Ala Lys Lys Tyr Glu
242 225 230 235 240
245 Gly Leu Arg Pro Glu Leu Lys Ser Ser Gly Met Arg Asp Ile Asp Ala
246 245 250 255
249 Thr Leu Thr Thr Arg Glu Leu Ala Tyr Met Ile Lys Lys Ala Gly Ile
250 260 265 270
253 Asp Phe Ala Lys Leu Pro Asp Gly Lys Arg Asp Ser Leu Met Gly Glu
254 275 280 285
257 Ser Thr Gly Gly Ala Thr Ile Phe Gly Val Thr Gly Gly Val Met Glu
258 290 295 300
261 Ala Ala Leu Arg Phe Ala Tyr Glu Ala Val Thr Gly Lys Lys Pro Asp
262 305 310 315 320
265 Ser Trp Asp Phe Lys Ala Val Arg Gly Leu Asp Gly Ile Lys Glu Ala
266 325 330 335
269 Thr Val Asn Val Gly Gly Thr Asp Val Lys Val Ala Val Val His Gly
270 340 345 350

```

RAW SEQUENCE LISTING

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:50

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

```

273 Ala Lys Arg Phe Lys Gln Val Cys Asp Asp Val Lys Ala Gly Lys Ser
274          355          360          365
277 Pro Tyr His Phe Ile Glu Tyr Met Ala Cys Pro Gly Gly Cys Val Cys
278          370          375          380
281 Gly Gly Gly Gln Pro Val Met Pro Gly Val Leu Glu Ala Met Asp Arg
282 385          390          395          400
285 Thr Thr Thr Arg Leu Tyr Ala Gly Leu Lys Lys Arg Leu Ala Met Ala
286          405          410          415
289 Ser Ala Asn Lys Ala
290          420
293 <210> SEQ ID NO: 3
294 <211> LENGTH: 468
295 <212> TYPE: PRT
296 <213> ORGANISM: Entamoeba histolytica
298 <400> SEQUENCE: 3
300 Met Pro Pro Lys Pro Ser His Thr Leu Thr Gly His Asp His Asn His
301 1          5          10          15
304 Ser Ile Gln Phe Asp Trp Ser Lys Cys Met Gly Cys Gly Met Cys Ala
305          20          25          30
308 Thr Lys Cys Thr Phe Gly Val Leu Val Lys Gln Pro Pro Lys Ile Pro
309          35          40          45
312 Pro Phe Val Gln Pro Asn Arg Glu Lys Leu Ser Gln Glu Asn Thr Asp
313          50          55          60
316 Lys Thr Arg Val Leu Ile Asp Glu Ser Glu Cys Thr Gly Cys Gly Gln
317 65          70          75          80
320 Cys Ser Leu Val Cys Asn Phe Gly Ser Ile Thr Pro Ile Asp His Leu
321          85          90          95
324 Val Asp Thr Phe Lys Ala Lys Glu Ala Gly Lys Lys Leu Val Ala Met
325          100          105          110
328 Ile Ala Pro Ser Thr Arg Leu Gly Val Ala Glu Ala Met Gly Met Pro
329          115          120          125
332 Ile Gly Ser Thr Ala Met Ala Gln Leu Val His Cys Leu Arg Leu Ile
333          130          135          140
336 Gly Phe Asp Tyr Val Phe Asp Val Asp Ala Gly Ala Asp Lys Thr Thr
337 145          150          155          160
340 Met Asp Asp Tyr Ala Glu Val Ile Glu Met Lys Lys Glu Gly Lys Gly
341          165          170          175
344 Pro Ala Ile Thr Ser Cys Cys Pro Ala Trp Ile Glu Leu Val Glu Lys
345          180          185          190
348 Glu Tyr Pro Asp Leu Ile Pro Asn Val Ser Thr Ala Arg Ser Pro Ile
349          195          200          205
352 Gly Cys Leu Ala Gly Cys Ile Lys Arg Gly Trp Ala Lys Asp Val Gly
353          210          215          220
356 Ile Ala Val Glu Asp Leu Tyr Thr Val Gly Ile Met Pro Cys Ile Ala
357 225          230          235          240
360 Lys Lys Thr Glu Ser Gln Arg Gln Gln Ile His Gln Asp Tyr Asp Ala
361          245          250          255
364 Ser Cys Thr Ser Asn Glu Ile Ala Ala Tyr Phe Lys Lys His Leu Pro
365          260          265          270

```

RAW SEQUENCE LISTING

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:50

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

```

368 Pro Glu Glu Cys Lys Phe Thr Gln Glu Arg Glu Glu Ala Leu Ala Lys
369      275      280      285
372 Thr Glu Asp Gly Gln Cys Asp Leu Pro Phe Arg Arg Ile Ser Gly Gly
373      290      295      300
376 Ser Asn Ile Phe Gly Lys Thr Gly Gly Val Cys Glu Thr Val Leu Arg
377 305      310      315      320
380 Val Ile Ala Arg Asn Ala Gly Val Asp Trp Asn Ser Cys Thr Val Asn
381      325      330      335
384 Lys Glu Glu Thr Phe Lys His Ala Ala Ser Gly Ser Thr Met Thr Asn
385      340      345      350
388 Leu Ser Val Asp Ile Gly Gly Thr Ile Ile Thr Gly Ala Val Cys His
389      355      360      365
392 Gly Gly Tyr Ala Ile Arg His Ala Cys Glu Leu Ile Arg Lys Gly Glu
393      370      375      380
396 Leu Lys Val Asp Val Val Glu Met Met Ala Cys Val Gly Gly Cys Leu
397 385      390      395      400
400 Gly Gly Ala Gly Gln Pro Lys Ile Pro Pro Ala Lys Lys Leu Glu Met
401      405      410      415
404 Asp Lys Arg Arg Val Met Leu Asp Ile Leu Asp Gln Gln Thr Asp Ile
405      420      425      430
408 Arg Ala Ala Asn Glu Asn Thr Asp Val Leu Gly Trp Ile Asp Lys His
409      435      440      445
412 Phe Asp His Gln Gly Ala His Gln His Leu His Thr Tyr Phe Thr Pro
413      450      455      460
416 Arg Tyr Gln Asn
417 465
420 <210> SEQ ID NO: 4
421 <211> LENGTH: 491
422 <212> TYPE: PRT
423 <213> ORGANISM: Saccharomyces cerevisiae
425 <400> SEQUENCE: 4
427 Met Ser Ala Leu Leu Ser Glu Ser Asp Leu Asn Asp Phe Ile Ser Pro
428 1      5      10      15
431 Ala Leu Ala Cys Val Lys Pro Thr Gln Val Ser Gly Gly Lys Lys Asp
432      20      25      30
435 Asn Val Asn Met Asn Gly Glu Tyr Glu Val Ser Thr Glu Pro Asp Gln
436      35      40      45
439 Leu Glu Lys Val Ser Ile Thr Leu Ser Asp Cys Leu Ala Cys Ser Gly
440      50      55      60
443 Cys Ile Thr Ser Ser Glu Glu Ile Leu Leu Ser Ser Gln Ser His Ser
444 65      70      75      80
447 Val Phe Leu Lys Asn Trp Gly Lys Leu Ser Gln Gln Gln Asp Lys Phe
448      85      90      95
451 Leu Val Val Ser Val Ser Pro Gln Cys Arg Leu Ser Leu Ala Gln Tyr
452      100      105      110
455 Tyr Gly Leu Thr Leu Glu Ala Ala Asp Leu Cys Leu Met Asn Phe Phe
456      115      120      125
459 Gln Lys His Phe Gln Cys Lys Tyr Met Val Gly Thr Glu Met Gly Arg
460      130      135      140

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/763,712C

DATE: 10/23/2006
TIME: 16:00:51

Input Set : E:\sequence listing for 10763712.txt
Output Set: N:\CRF4\10232006\J763712C.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:82; Xaa Pos. 85,124

Seq#:185; Xaa Pos. 1,2,3,5,6,7,15

Seq#:186; Xaa Pos. 3,6

VERIFICATION SUMMARY

DATE: 10/23/2006

PATENT APPLICATION: US/10/763,712C

TIME: 16:00:51

Input Set : E:\sequence listing for 10763712.txt

Output Set: N:\CRF4\10232006\J763712C.raw

L:9996 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:82 after pos.:80

M:341 Repeated in SeqNo=82

L:16854 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:185 after pos.:0

L:16879 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:186 after pos.:0